

Dead Code Elimination Based Pointer Analysis for Multithreaded Programs

Mohamed A. El-Zawawy

College of Computer and Information Sciences
Al-Imam M. I.-S. I. University
Riyadh 11432
Kingdom of Saudi Arabia
and
Department of Mathematics
Faculty of Science
Cairo University
Giza 12613
Egypt
maelzawawy@cu.edu.eg

Abstract. This paper presents a new approach for optimizing multithreaded programs with pointer constructs. The approach has applications in the area of certified code (proof-carrying code) where a justification or a proof for the correctness of each optimization is required. The optimization meant here is that of dead code elimination.

Towards optimizing multithreaded programs the paper presents a new operational semantics for parallel constructs like join-fork constructs, parallel loops, and conditionally spawned threads. The paper also presents a novel type system for flow-sensitive pointer analysis of multithreaded programs. This type system is extended to obtain a new type system for live-variables analysis of multithreaded programs. The live-variables type system is extended to build the third novel type system, proposed in this paper, which carries the optimization of dead code elimination. The justification mentioned above takes the form of type derivation in our approach.

1 Introduction

One of the mainstream programming approaches today is multithreading. Using multiple threads is useful in many ways like (a) concealing suspension caused by some commands, (b) making it easier to build huge software systems, (c) improving execution of programs especially those that are executed on multiprocessors, and (d) building advanced user interfaces.

Short title: Dead Code Elimination for Multithreaded Programs.
Mathematics Subject Classification: 68Q55, 68N19.

The potential interaction between threads in a multithreaded program complicates both the compilation and the program analysis processes. Moreover this interaction also makes it difficult to extend the scope of program analysis techniques of sequential programs to cover multithreaded programs.

Typically optimizing multithreaded programs is achieved in an algorithmic form using data-flow analyses. This includes transforming the given program into a control-flow graph which is a convenient form for the algorithm to manipulate. For some applications like certified code, it is desirable to associate each program optimization with a justification or a proof for the correctness of the optimization. For these cases, the algorithmic approach to program analysis is not a good choice as it does not work on the syntactical structure of the program and hence does not reflect the transformation process. Moreover the desired justification must be relatively simple as it gets checked within a trusted computing base.

Type systems stand as a convenient alternative for the algorithmic approach of program analyses when a justification is necessary. In the type systems approach, analysis and optimization of programs are directed by the syntactical structure of the program. Inference rules of type systems are advantageously relatively simple and so is the justification which takes the form of a type derivation in this case. The adequacy of type systems for program analysis has already been studied for example in [1,2,3]

Pointer analysis is among the most important program analyses and it calculates information describing contents of pointers at different program points. The application of pointer analysis to multithreaded programs results in information that is required for program analyses and compiler optimizations such as live-variables analysis and dead code elimination, respectively. The live-variables analysis finds for each program point the set of variables whose values are used usefully in the rest of the program. The results of live-variables analysis is necessary for the optimization of dead code elimination which removes code that has no effect on values of variables of interest at the end of the program.

This paper presents a new approach for optimizing multithreaded programs with pointer constructs. The scope of the proposed approach is broad enough to include certified (proof-carrying) code applications where a justification for optimization is necessary. Type systems are basic tools of the new approach which considers structured parallel constructs such as join-fork constructs, parallel loops, and conditionally spawned threads. The justifications in our approach take the form of type derivations. More precisely, the paper presents a type system for flow-sensitive pointer analysis of multithreaded programs. The live-variables analysis of multithreaded programs is also treated in this paper by a type system which is an extension of the type system for pointer analysis. The extension has the form of another component being added to points-to types. The dead code elimination of multithreaded programs is then achieved using a type system which is again an extension of the type system for live-variables analysis. This time the extension takes the form of a transformation component

added to inference rules of the type system for live-variables analysis. To prove the soundness of the three proposed type systems, a novel operational semantics for parallel constructs is proposed in this paper.

<pre> 1. x := &y; 2. * x := 2; 3. par{ 4. {y := 4} 5. {y := 5; 6. * x := 6} 7. }; 8. x := 8; 9. x := 9 </pre>	\Rightarrow	<pre> x := &y; skip; par{ {y := 4} {skip; * x := 6} }; skip; x := 9 </pre>
--	---------------	--

Fig. 1. A motivating example

Some comments are in order. The language that we use does not allow pointer arithmetic, so it can be thought of as a Pascal-style language. The parallel construct assumes that each thread is executed atomically. This is a restriction that is not very common to many implementation of threads. But based on our construct, it should be easy to treat a more general construct. Writing computer programs to efficiently calculate the types, using our type systems, is not hard. This is so because each time the program calculates a type, the program will be given a pre-type and seeking its post type or vice versa. In the semantics our language, we assume that \mathbb{Z} and the set of addresses, *Addr*s, are disjoint sets.

Motivation

Figure 1 presents a motivating example of the work presented in this paper. Consider the program on the left-hand-side of the figure. Suppose that at the end of the program we are interested in the values of x and y . We note that the assignment in line 8 is dead code as the variable x is modified in line 9 before we make any use of the value that the variable gets in line 8. The assignment in line 2 indirectly modifies y which is modified again in the *par* command before any useful use of the value that y gets in line 2. Therefore line 2 is dead code. The *par* command has two threads which can be executed in any order. If the first thread is executed first then assignments in lines 4 and 5 become dead code. If the second thread is executed first then assignments in lines 5 and 6 become dead code. Therefore the dead code in the *par* command is the assignment in line 5 only.

This paper presents a technique that discovers and removes such dead code in parallel structured programs with pointer constructs. The output of the technique is a program like that on the right-hand-side of Figure 1. In addition to the join-fork construct (*par*), the paper also considers other parallel constructs like conditionally spawned threads and parallel loops. With each such program

optimization, our technique presents a justification or a proof for the correctness of the optimization. The proof takes the form of a type derivation.

Contributions

Contributions of this paper are the following:

1. A simple yet powerful operational semantics for multi-threaded programs with pointer constructs.
2. A novel type system for pointer analysis of multithreaded programs. To our knowledge, this is the first attempt to use type systems for pointer analysis of multithreaded programs.
3. A new type systems for live-variables analysis of multithreaded programs.
4. An original type system for the optimization of dead code elimination for multithreaded programs.

Organization The rest of the paper is organized as follows. The language that we study (a while language enriched with pointer and parallel constructs) and an operational semantics for its constructs are presented in Section 2. Sections 3 and 4 present our type systems for flow-sensitive pointer and live-variables analyses, respectively. The type system carrying program optimization is introduced in Section 5. Related work is discussed in Section 6.

2 Programming language

This section presents the programming language (Figure 2) we use together with an operational semantics for its constructs. The language is the simple *while* language [4] enriched with commands for pointer manipulations and structured parallel constructs.

$$\begin{aligned}
& n \in \mathbb{Z}, x \in \text{Var}, \text{ and } \oplus \in \{+, -, \times\} \\
& e \in \text{Aexprs} ::= x \mid n \mid e_1 \oplus e_2 \\
& b \in \text{Bexprs} ::= \text{true} \mid \text{false} \mid \neg b \mid e_1 = e_2 \mid e_1 \leq e_2 \mid b_1 \wedge b_2 \mid b_1 \vee b_2 \\
& S \in \text{Stmts} ::= x := e \mid x := \&y \mid *x := e \mid x := *y \mid \text{skip} \mid S_1; S_2 \mid \text{if } b \text{ then } S_t \text{ else } S_f \mid \\
& \quad \text{while } b \text{ do } S_t \mid \text{par}\{\{S_1\}, \dots, \{S_n\}\} \mid \text{par-if}\{(b_1, S_1), \dots, (b_n, S_n)\} \mid \text{par-for}\{S\}.
\end{aligned}$$

Fig. 2. The programming language.

The parallel constructs include join-fork constructs, parallel loops, and conditionally spawned threads. The *par* (join-fork) construct starts executing many concurrent threads at the beginning of the *par* construct and then waits until the

completion of all these executions at the end of the *par* construct. Semantically, the *par* construct can be expressed approximately as if the threads are executed sequentially in an arbitrary order. The parallel loop construct included in our language is that of *par-for*. This construct executes, in parallel, a statically unknown number of threads each of which has the same code (the loop body). Therefore the semantics of *par-for* can be expressed using that of the *par* construct. The construct including conditionally spawned threads is that of *par-if*. This construct executes, in parallel, its n concurrent threads. The execution of thread (b_i, S_i) includes the execution of S_i only if b_i is true.

One way to define the meaning of the constructs of our programming language, including the parallel constructs, is by an operational semantics. This amounts to defining a transition relation \rightsquigarrow between states which are defined as follows.

Definition 1. 1. $Addr s = \{x' \mid x \in Var\}$ and $Val = \mathbb{Z} \cup Addr s$.
2. $state \in States = \{abort\} \cup \{\gamma \mid \gamma \in \Gamma = Var \longrightarrow Val\}$.

The semantics of arithmetic and Boolean expressions are defined as usual except that arithmetic and Boolean operations are not allowed on pointers.

$$\begin{aligned} \llbracket n \rrbracket \gamma &= n & \llbracket \&x \rrbracket \gamma &= x' & \llbracket x \rrbracket \gamma &= \gamma(x) & \llbracket true \rrbracket \gamma &= true & \llbracket false \rrbracket \gamma &= false \\ \llbracket *x \rrbracket \gamma &= \begin{cases} \gamma(y) & \text{if } \gamma(x) = y', \\ ! & \text{otherwise.} \end{cases} & \llbracket e_1 \oplus e_2 \rrbracket \gamma &= \begin{cases} \llbracket e_1 \rrbracket \gamma \oplus \llbracket e_2 \rrbracket \gamma & \text{if } \llbracket e_1 \rrbracket \gamma, \llbracket e_2 \rrbracket \gamma \in \mathbb{Z}, \\ ! & \text{otherwise.} \end{cases} \\ \llbracket \neg A \rrbracket \gamma &= \begin{cases} \neg(\llbracket A \rrbracket \gamma) & \text{if } \llbracket A \rrbracket \gamma \in \{true, false\}, \\ ! & \text{otherwise.} \end{cases} & \llbracket e_1 = e_2 \rrbracket \gamma &= \begin{cases} ! & \text{if } \llbracket e_1 \rrbracket \gamma = ! \text{ or } \llbracket e_2 \rrbracket \gamma = !, \\ true & \text{if } \llbracket e_1 \rrbracket \gamma = \llbracket e_2 \rrbracket \gamma \neq !, \\ false & \text{otherwise.} \end{cases} \\ \llbracket e_1 \leq e_2 \rrbracket \gamma &= \begin{cases} ! & \text{if } \llbracket e_1 \rrbracket \gamma \notin \mathbb{Z} \text{ or } \llbracket e_2 \rrbracket \gamma \notin \mathbb{Z}, \\ \llbracket e_1 \rrbracket \gamma \leq \llbracket e_2 \rrbracket \gamma & \text{otherwise.} \end{cases} \\ \text{For } \diamond \in \{\wedge, \vee\}, \llbracket b_1 \diamond b_2 \rrbracket \gamma &= \begin{cases} ! & \text{if } \llbracket b_1 \rrbracket \gamma = ! \text{ or } \llbracket b_2 \rrbracket \gamma = !, \\ \llbracket b_1 \rrbracket \gamma \diamond \llbracket b_2 \rrbracket \gamma & \text{otherwise.} \end{cases} \end{aligned}$$

The inference rules of our semantics (transition relation) are defined as follows:

$$\begin{array}{c} \frac{\llbracket e \rrbracket \gamma = !}{x := e : \gamma \rightsquigarrow abort} \quad \frac{\llbracket e \rrbracket \gamma \neq !}{x := e : \gamma \rightsquigarrow \gamma[x \mapsto \llbracket e \rrbracket \gamma]} \quad \frac{\gamma(x) = z' \quad z := e : \gamma \rightsquigarrow state}{*x := e : \gamma \rightsquigarrow state} \\ \frac{\gamma(x) \notin Addr s}{*x := e : \gamma \rightsquigarrow abort} \quad \frac{}{x := \&y : \gamma \rightsquigarrow \gamma[x \mapsto y']} \quad \frac{\gamma(y) = z' \quad x := z : \gamma \rightsquigarrow \gamma'}{x := *y : \gamma \rightsquigarrow \gamma'} \\ \frac{\gamma(y) \notin Addr s}{x := *y : \gamma \rightsquigarrow abort} \quad \frac{}{skip : \gamma \rightsquigarrow \gamma} \quad \frac{S_1 : \gamma \rightsquigarrow abort}{S_1; S_2 : \gamma \rightsquigarrow abort} \quad \frac{S_1 : \gamma \rightsquigarrow \gamma'' \quad S_2 : \gamma'' \rightsquigarrow state}{S_1; S_2 : \gamma \rightsquigarrow state} \\ \frac{\llbracket b \rrbracket \gamma = !}{if \ b \ \text{then} \ S_t \ \text{else} \ S_f : \gamma \rightsquigarrow abort} \quad \frac{\llbracket b \rrbracket \gamma = true \quad S_t : \gamma \rightsquigarrow state}{if \ b \ \text{then} \ S_t \ \text{else} \ S_f : \gamma \rightsquigarrow state} \\ \frac{\llbracket b \rrbracket \gamma = false \quad S_f : \gamma \rightsquigarrow state}{if \ b \ \text{then} \ S_t \ \text{else} \ S_f : \gamma \rightsquigarrow state} \quad \frac{\llbracket b \rrbracket \gamma = !}{while \ b \ \text{do} \ S_t : \gamma \rightsquigarrow abort} \quad \frac{\llbracket b \rrbracket \gamma = false}{while \ b \ \text{do} \ S_t : \gamma \rightsquigarrow \gamma} \\ \frac{\llbracket b \rrbracket \gamma = true \quad S : \gamma \rightsquigarrow \gamma'' \quad while \ b \ \text{do} \ S_t : \gamma'' \rightsquigarrow state}{while \ b \ \text{do} \ S_t : \gamma \rightsquigarrow state} \quad \frac{\llbracket b \rrbracket \gamma = true \quad S : \gamma \rightsquigarrow abort}{while \ b \ \text{do} \ S_t : \gamma \rightsquigarrow abort} \end{array}$$

- **Join-fork:**

$$\frac{}{\text{par}\{\{S_1\}, \dots, \{S_n\}\} : \gamma \rightsquigarrow \gamma'}^\dagger \quad \frac{}{\text{par}\{\{S_1\}, \dots, \{S_n\}\} : \gamma \rightsquigarrow \text{abort}}^\ddagger$$

\dagger there exist a permutation $\theta : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ and $n + 1$ states $\gamma = \gamma_1, \dots, \gamma_{n+1} = \gamma'$ such that for every $1 \leq i \leq n$, $S_{\theta(i)} : \gamma_i \rightarrow \gamma_{i+1}$.

\ddagger there exist m such that $1 \leq m \leq n$, a one-to-one map $\beta : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$, and $m + 1$ states $\gamma = \gamma_1, \dots, \gamma_{m+1} = \text{abort}$ such that for every $1 \leq i \leq m$, $S_{\beta(i)} : \gamma_i \rightarrow \gamma_{i+1}$.

- **Conditionally spawned threads:**

$$\frac{\text{par}\{\{\text{if } b_1 \text{ then } S_1 \text{ else skip}\}, \dots, \{\text{if } b_n \text{ then } S_n \text{ else skip}\}\} : \gamma \rightsquigarrow \gamma'}{\text{par-if}\{(b_1, S_1), \dots, (b_n, S_n)\} : \gamma \rightsquigarrow \gamma'}$$

$$\frac{\text{par}\{\{\text{if } b_1 \text{ then } S_1 \text{ else skip}\}, \dots, \{\text{if } b_n \text{ then } S_n \text{ else skip}\}\} : \gamma \rightsquigarrow \text{abort}}{\text{par-if}\{(b_1, S_1), \dots, (b_n, S_n)\} : \gamma \rightsquigarrow \text{abort}}$$

- **Parallel loops:**

$$\frac{\exists n. \overbrace{\text{par}\{\{S\}, \dots, \{S\}\}}^{n\text{-times}} : \gamma \rightsquigarrow \gamma'}{\text{par-for}\{S\} : \gamma \rightsquigarrow \gamma'}$$

$$\frac{\exists n. \overbrace{\text{par}\{\{S\}, \dots, \{S\}\}}^{n\text{-times}} : \gamma \rightsquigarrow \text{abort}}{\text{par-for}\{S\} : \gamma \rightsquigarrow \text{abort}}$$

A simple example for the *par-for* command is *par-for*{ $x := 10$ }. The execution of this command amounts to fixing a number randomly, say 7, and then to concurrently execute the seven threads $\{x := 10\}_1, \dots, \{x := 10\}_7$.

3 Pointer analysis

In this section, we present a novel technique for flow-sensitive pointer analysis of structured parallel programs where shared pointers may be updated simultaneously. Our technique manipulates important parallel constructs; join-fork constructs, parallel loops, and conditionally spawned threads. The proposed technique has the form of a compositional type system which is simply structured. Consequently results of the analysis are in the form of types assigned to expressions and statements approved by type derivations. Therefore a type is assigned to each program point of a statement (program). This assigned type specifies for each variable in the program a conservative approximation of the addresses that may be assigned to the variable. The set of points-to types *PTS* and the relation $\models \subseteq \Gamma \times \text{PTS}$ are defined as follows:

Definition 2. 1. $\text{PTS} = \{\text{pts} \mid \text{pts} : \text{Var} \rightarrow 2^{\text{Addr}}\}$.

2. $\text{pts} \leq \text{pts}' \stackrel{\text{def}}{\iff} \forall x \in \text{Var}. \text{pts}(x) \subseteq \text{pts}'(x)$.

3. $\gamma \models \text{pts} \stackrel{\text{def}}{\iff} (\forall x \in \text{Var}. \gamma(x) \in \text{Addr} \implies \gamma(x) \in \text{pts}(x))$.

The judgement of an expression has the form $e : pts \rightarrow A$. The intended meaning of this judgment, which is formalized in Lemma 1, is that A is the collection of addresses that e may evaluate to in a state of type pts . The judgement of a statement has the form $S : pts \rightarrow pts'$. This judgement simply guarantees that if S is executed in a state of type pts and the execution terminates in a state γ' , then γ' has type pts' . Typically, the pointer analysis for a program S is achieved via a post-type derivation for the bottom type (mapping variables to \emptyset) as the pre-type.

The inference rules of our type system for pointer analysis are the following:

$$\begin{array}{c}
\frac{}{n : pts \rightarrow \emptyset} \quad \frac{}{x : pts \rightarrow pts(x)} \quad \frac{}{e_1 \oplus e_2 : pts \rightarrow \emptyset} \quad \frac{e : pts \rightarrow A}{x := e : pts \rightarrow pts[x \mapsto A]} (:=^p) \\
\\
\frac{}{x := \&y : pts \rightarrow pts[x \mapsto \{y'\}]} (:= \&^p) \quad \frac{}{skip : pts \rightarrow pts} \\
\frac{\forall z' \in pts(y). x := z : pts \rightarrow pts'}{x := *y : pts \rightarrow pts'} (:= *^p) \quad \frac{\forall z' \in pts(x). z := e : pts \rightarrow pts'}{*x := e : pts \rightarrow pts'} (* :=^p) \\
\frac{S_i : pts \cup \cup_{j \neq i} pts_j \rightarrow pts_i}{par\{\{S_1\}, \dots, \{S_n\}\} : pts \rightarrow \cup_i pts_i} (par^p) \quad \frac{S_1 : pts \rightarrow pts'' \quad S_2 : pts'' \rightarrow pts'}{S_1; S_2 : pts \rightarrow pts'} (seq^p) \\
\frac{par\{\{if\ b_1\ then\ S_1\ else\ skip\}, \dots, \{if\ b_n\ then\ S_n\ else\ skip\}\} : pts \rightarrow pts'}{par-if\{(b_1, S_1), \dots, (b_n, S_n)\} : pts \rightarrow pts'} (par-if^p) \\
\frac{S : pts \cup pts' \rightarrow pts'}{par-for\{S\} : pts \rightarrow pts'} (par-for^p) \quad \frac{S_t : pts \rightarrow pts' \quad S_f : pts \rightarrow pts'}{if\ b\ then\ S_t\ else\ S_f : pts \rightarrow pts'} (if^p) \\
\frac{S_i : pts \rightarrow pts}{while\ b\ do\ S_i : pts \rightarrow pts} (whl^p) \quad \frac{pts'_1 \leq pts_1 \quad S : pts_1 \rightarrow pts_2 \quad pts_2 \leq pts'_2}{S : pts'_1 \rightarrow pts'_2} (csq^p)
\end{array}$$

The inference rules corresponding to assignment commands are clear. For the rule (par^p) of the join-fork command, par , one possibility is that the execution of a thread S_i starts before the execution of any other thread starts. Another possibility is that the execution starts after executions of all other threads end. Of course there are many other possibilities in between. Consequently, the analysis of the thread S_i must consider all such possibilities. This is reflected in the pre-type of S_i and the post-type of the par command. Clearly pts , in (par^p) , is the given pre-type for the par command and for which the rule calculates a post-type. The union operation in the rule (par^p) makes the pre-type of threads general enough. Similar explanations clarify the rules $(par-if^p)$ and $(par-for^p)$.

We note that a type invariant is required to type a *while* statement. Also to achieve the analysis for one of the par 's threads we need to know the analysis results for all other threads. However obtaining these results requires the result of analyzing the first thread. Therefore there is a kind of circularity in rule (par^p) . Similar situations are in rules $(par-if^p)$ and $(par-for^p)$. Such issues can be treated using a fix-point algorithm. The convergence of this algorithm is guaranteed as the rules of our type system are monotone and the set of points-to types PTS is a complete lattice. What makes calculations actual simple is that for any given

program the lattice PTS is finite. The rule (csq^p) is necessary to calculate a type invariant.

Lemma 1. 1. Suppose $e : pts \rightarrow A$ and $\gamma \models pts$. Then $\llbracket e \rrbracket \gamma \in Addr$ implies $\llbracket e \rrbracket \gamma \in A$.
 2. $pts \leq pts' \iff (\forall \gamma. \gamma \models pts \implies \gamma \models pts')$.

Proof. The first item is obvious. The left-to-right direction of (2) is easy. The other direction is proved as follows. Suppose $y' \in pts(x)$. Then the state $\{(x, y'), (t, 0) \mid t \in Var \setminus \{x\}\}$ is of type pts and hence of type pts' implying that $y' \in pts'(x)$. Therefore $pts(x) \subseteq pts'(x)$. Since x is arbitrary, $pts \leq pts'$.

Theorem 1. (Soundness) Suppose that $S : pts \rightarrow pts'$, $S : \gamma \rightsquigarrow \gamma'$, and $\gamma \models pts$. Then $\gamma' \models pts'$.

Proof. The proof is by structural induction on the type derivation. We demonstrate some cases.

- The case of $(:=^p)$: In this case $pts' = pts[x \mapsto A]$ and $\gamma' = \gamma[x \mapsto \llbracket e \rrbracket \gamma]$. Therefore by the previous lemma $\gamma \models pts$ implies $\gamma' \models pts'$.
- The case of $(* :=^p)$: In this case there exists $z \in Var$ such that $\gamma(x) = z'$ and $z := e : \gamma \rightsquigarrow \gamma'$. Because $\gamma \models pts$, $z' \in pts(x)$ and hence by assumption $z := e : pts \rightarrow pts'$. Therefore by soundness of $(:=^p)$, $\gamma' \models pts'$.
- The case of (par^p) : In this case there exist a permutation $\theta : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ and $n + 1$ states $\gamma = \gamma_1, \dots, \gamma_{n+1} = \gamma'$ such that for every $1 \leq i \leq n$, $S_{\theta(i)} : \gamma_i \rightarrow \gamma_{i+1}$. Also $\gamma_1 \models pts$ implies $\gamma_1 \models pts \cup \bigcup_{j \neq \theta(1)} pts_j$. Therefore by the induction hypothesis $\gamma_2 \models pts_{\theta(1)}$. This implies $\gamma_2 \models pts \cup \bigcup_{j \neq \theta(2)} pts_j$. Again by the induction hypothesis we get $\gamma_3 \models pts_{\theta(2)}$. Therefore by a simple induction on n , we can show that $\gamma' = \gamma_{n+1} \models pts_{\theta(n)}$ which implies $\gamma' \models pts' = \bigcup_j pts_j$.
- The case of $(par - for^p)$: In this case there exists n such that $\overbrace{par\{\{S\}, \dots, \{S\}\}}^{n\text{-times}} : \gamma \rightsquigarrow \gamma'$. By induction hypothesis we have $S : pts \cup pts' \rightarrow pts'$. By (par^p) we conclude that $\overbrace{par\{\{S\}, \dots, \{S\}\}}^{n\text{-times}} : pts \rightsquigarrow pts'$. Therefore by the soundness of (par^p) , $\gamma' \models pts'$.

4 Live-variables analysis

In this section, we present a type system to perform live-variables analysis for pointer programs with structured parallel constructs. We start with defining live-variables:

Definition 3. – A variable is usefully used if it is used

- as the operand of the unary operation $*$.
- in an assignment to a variable that is live at the end of the assignment, or

- in the guard of an if-statement or a while-statement,
- A variable is live at a program point if there is a computational path from that program point during which the variable gets usefully used before being modified.

Definition 4. The set of live types is denoted by L and equal to $PTS \times \mathcal{P}(\text{Var})$. The second component of a live type is termed a live-component. The subtyping relation \leq is defined as: $(pts, l) \leq (pts', l') \stackrel{\text{def}}{\iff} pts \leq pts' \text{ and } l \supseteq l'$.

The live-variables analysis is a backward analysis. For each program point, this analysis specifies the set of variables that may be live (according to the definition above) at that point.

Our type system for live-variables analysis is obtained as an enrichment of the type system for pointer analysis, presented in the previous section. Hence one can say that the type system presented here is a strict extension of that presented above. This is so because the result of pointer analysis is necessary to improve the precision of the live-variables analysis. This also gives an intuitive explanation of the definition of live types above.

The judgement of a statement S has the form $S : (pts, l) \rightarrow (pts', l')$. The intuition of the judgement is that the presence of live-variables at the post-state of an execution of S in l' implies the presence of live-variables at the pre-state of this execution in l . The intuition agrees with the fact that live-variables analysis is a backward analysis and gives an insight into the definition of $\gamma \models l$ below.

Suppose we have the set of variables l' that we have interest in their values at the end of executing a statement S and the result of pointer analysis of S (in the form $S : pts \rightarrow pts'$). The live-variables analysis takes the form of a pre-type derivation that calculates a set l such that $S : (pts, l) \rightarrow (pts', l')$. The idea here is that by proceeding from the last point of the program, calculating pre-types, we achieve the backward analysis.

The inference rules for our type system for live-variables analysis are as follows.

$$\begin{array}{c}
\frac{x := e : pts \rightarrow pts' \quad x \notin l'}{x := e : (pts, l') \rightarrow (pts', l')} (:=_1^l) \quad \frac{x := e : pts \rightarrow pts' \quad x \in l'}{x := e : (pts, (l' \setminus \{x\}) \cup FV(e)) \rightarrow (pts', l')} (:=_2^l) \\
\\
\frac{}{x := \&y : (pts, l' \setminus \{x\}) \rightarrow (pts[x \mapsto \{y\}], l')} (:= \&^l) \quad \frac{}{skip : (pts, l) \rightarrow (pts, l)} \\
\\
\frac{x := *y : pts \rightarrow pts' \quad x \notin l'}{x := *y : (pts, l' \cup \{y\}) \rightarrow (pts', l')} (:= *_1^l) \\
\\
\frac{x := *y : pts \rightarrow pts' \quad x \in l'}{x := *y : (pts, (l' \setminus \{x\}) \cup \{y, z \mid z' \in pts(y)\}) \rightarrow (pts', l')} (:= *_2^l) \\
\\
\frac{*x := e : pts \rightarrow pts' \quad pts(x) \cap l' = \emptyset}{*x := e : (pts, l' \cup \{x\}) \rightarrow (pts', l')} (* :=_1^l) \quad \frac{*x := e : pts \rightarrow pts' \quad pts(x) \cap l' \neq \emptyset}{*x := e : (pts, l' \cup FV(e) \cup \{x\}) \rightarrow (pts', l')} (* :=_2^l) \\
\\
\frac{S_i : (pts \cup \cup_{j \neq i} pts_j, l_i) \rightarrow (pts_i, l' \cup \cup_{j \neq i} l_j)}{par\{\{S_1\}, \dots, \{S_n\}\} : (pts, \cup_i l_i) \rightarrow (\cup_i pts_i, l')} (par^l)
\end{array}$$

$$\begin{array}{c}
\frac{\text{par}\{\{\text{if } b_1 \text{ then } S_1 \text{ else skip}\}, \dots, \{\text{if } b_n \text{ then } S_n \text{ else skip}\}\} : (pts, l) \rightarrow (pts', l')}{\text{par-if}\{(b_1, S_1), \dots, (b_n, S_n)\} : (pts, l) \rightarrow (pts', l')} \text{ (par-if}^l\text{)} \\
\frac{S : (pts \cup pts', l) \rightarrow (pts', l' \cup l)}{\text{par-for}\{S\} : (pts, l) \rightarrow (pts', l')} \text{ (par-for}^l\text{)} \\
\frac{S_1 : (pts, l) \rightarrow (pts'', l'') \quad S_2 : (pts'', l'') \rightarrow (pts', l')}{S_1; S_2 : (pts, l) \rightarrow (pts', l')} \text{ (seq}^l\text{)} \\
\frac{S_t, S_f : (pts, l) \rightarrow (pts', l')}{\text{if } b \text{ then } S_t \text{ else } S_f : (pts, l \cup FV(b)) \rightarrow (pts', l')} \text{ (if}^l\text{)} \\
\frac{l = l' \cup FV(b) \quad S_t : (pts, l') \rightarrow (pts, l)}{\text{while } b \text{ do } S_t : (pts, l) \rightarrow (pts, l')} \text{ (whl}^l\text{)} \\
\frac{(pts'_1, l'_1) \leq (pts_1, l_1) \quad S : (pts_1, l_1) \rightarrow (pts_2, l_2) \quad (pts_2, l_2) \leq (pts'_2, l'_2)}{S : (pts'_1, l'_1) \rightarrow (pts'_2, l'_2)} \text{ (csq}^l\text{)}
\end{array}$$

For the command $*x := e$, we have two rules, namely $(* :=_1^l)$ and $(* :=_2^l)$. In both cases, calculating the pre-type from the post-type includes adding x to the post-type. This is so because according to Definition 3, x is live at the pre-state of any execution of the command. The rule $(* :=_1^l)$ deals with the case that there is no possibility that the variable modified by this statement is live ($pts(x) \cap l' = \emptyset$) at the end of an execution. In this case there is no need to add any other variables to the post-type. The rule $(* :=_2^l)$ deals with the case that there is a possibility that the variable modified by this statement is live ($pts(x) \cap l' \neq \emptyset$) at the end of an execution. In this case, there is a possibility that free variables of e are used usefully according to Definition 3. Therefore free variables of e are added to the post-type. This gives an intuitive explanation for rules of all the assignment commands. The intuition given in the previous section for the rules (par^l) , $(par\text{-if}^l)$, and $(par\text{-for}^l)$.

Towards proving the soundness of our type system for live-variables analysis, we introduce necessary definitions and results.

Definition 5. 1. $\gamma \models_l pts \stackrel{\text{def}}{\iff} \forall x \in l. \gamma(x) \in \text{Addrs} \implies \gamma(x) \in pts(x)$.
2. $\gamma \sim_l \gamma' \stackrel{\text{def}}{\iff} \forall x \in l. \gamma(x) = \gamma'(x)$.
3. $\gamma \sim_{(pts, l)} \gamma' \stackrel{\text{def}}{\iff} \gamma \models_l pts, \gamma' \models_l pts$, and $\gamma \sim_l \gamma'$.

Definition 6. The expression $\gamma \models l$ denotes the case when there is a variable that is live at that state (computational point) and is not included in l . A state γ has type (pts, l) , denoted by $\gamma \models (pts, l)$, if $\gamma \models_l pts$ and $\gamma \models l$.

The following lemma is proved by structural induction on e and b .

Lemma 2. Suppose that γ and γ' are states and l and $l' \in \mathcal{P}(\text{Var})$. Then

1. If $l \supseteq l'$ and $\gamma \sim_l \gamma'$, then $\gamma \sim_{l'} \gamma'$.
2. If $l = l' \cup FV(e)$ and $\gamma \sim_l \gamma'$, then $\llbracket e \rrbracket \gamma = \llbracket e \rrbracket \gamma'$ and $\gamma \sim_{l'} \gamma'$.
3. If $l = l' \cup FV(b)$ and $\gamma \sim_l \gamma'$, then $\llbracket b \rrbracket \gamma = \llbracket b \rrbracket \gamma'$ and $\gamma \sim_{l'} \gamma'$.

The following lemma follows from Lemma 1.

Lemma 3. Suppose that $\gamma \models_l pts, FV(e) \subseteq l$, and $e : pts \rightarrow A$. Then

$$\llbracket e \rrbracket \gamma \in Addr s \implies \llbracket e \rrbracket \gamma \in A.$$

Proof. Consider the state γ' , where $\gamma' = \lambda x. \text{ if } x \in FV(e) \text{ then } \gamma(x) \text{ else } 0$. It is not hard to see that $\llbracket e \rrbracket \gamma = \llbracket e \rrbracket \gamma'$ and $\gamma' \models pts$. Now by Lemma 1, $\llbracket e \rrbracket \gamma' \in Addr s$ implies $\llbracket e \rrbracket \gamma' \in A$ which completes the proof.

Theorem 2. 1. $(pts, l) \leq (pts', l') \implies (\forall \gamma. \gamma \models_l pts \implies \gamma \models_{l'} pts')$.
 2. Suppose that $S : (pts, l) \rightarrow (pts', l')$ and $S : \gamma \rightsquigarrow \gamma'$. Then $\gamma \models_l pts$ implies $\gamma' \models_{l'} pts'$.
 3. Suppose that $S : (pts, l) \rightarrow (pts', l')$ and $S : \gamma \rightsquigarrow \gamma'$. Then $\gamma \models l$ implies $\gamma' \models l'$. This guarantees that if the set of variables live at γ' is included in l' , then the set of variables live at γ is included in l .

Proof. 1. Suppose $\gamma \models_l pts$. This implies $\gamma \models_{l'} pts$ because $l' \subseteq l$. The last fact implies $\gamma \models_{l'} pts'$ because $pts \leq pts'$.

2. The proof is by induction on the structure of type derivation. We show some cases.

- (a) The type derivation has the form $(:=_1^l)$. In this case, $pts' = pts[x \mapsto A]$ and $\gamma' = \gamma[x \mapsto \llbracket e \rrbracket \gamma]$. Therefore $\gamma \models_{l'} pts$ implies $\gamma' \models_{l'} pts'$ because $x \notin l'$.
- (b) The type derivation has the form $(:=_2^l)$. In this case, $e : pts \rightarrow A$, $pts' = pts[x \mapsto A]$, $\gamma' = \gamma[x \mapsto \llbracket e \rrbracket \gamma]$, and $l = (l' \setminus \{x\}) \cup FV(e)$. Therefore by Lemma 3 it is not hard to see $\gamma' \models_{l'} pts'$.
- (c) The type derivation has the form $(:=_1^{*l})$. In this case, for every $z' \in pts(y)$, we have $x := z : pts \rightarrow pts', \gamma(y) = z'$, and $x := z : \gamma \rightarrow \gamma'$. We have $z' \in pts(y)$, because $y \in l$ and $\gamma \models_l pts$. Therefore by $(:=_1^l)$, we have $x := z : (pts, l') \rightarrow (pts', l')$. Now $\gamma \models_l pts$ amounts to $\gamma \models_{l'} pts$. Hence we get $\gamma' \models_{l'} pts'$ by soundness of $(:=_1^l)$.
- (d) The type derivation has the form $(:=_2^{*l})$. In this case, for every $z' \in pts(y)$, we have $x := z : pts \rightarrow pts', \gamma(y) = z'$, $x := z : \gamma \rightarrow \gamma'$, and $l = (l' \setminus \{x\}) \cup \{y, z \mid z' \in pts(y)\}$. We have $z \in pts(y)$ because $\gamma \models_l pts$ and $y \in l$. Therefore by $(:=_2^l)$ we have $x := z : (pts, (l' \setminus \{x\}) \cup \{z\}) \rightarrow (pts', l')$. $\gamma \models_l pts$ implies $\gamma \models_{(l' \setminus \{x\}) \cup \{z\}} pts$. Hence by soundness of $(:=_2^l)$, we get $\gamma' \models_{l'} pts'$.
- (e) The type derivation has the form $(*_1^l)$. In this case, for every $z' \in pts(x)$, we have $z := e : pts \rightarrow pts', \gamma(x) = z'$, and $z := e : \gamma \rightarrow \gamma'$. We have $z' \in pts(x)$, because $x \in l$ and $\gamma \models_l pts$. Therefore by $(:=_1^l)$, we have $z := e : (pts, l') \rightarrow (pts', l')$ because $pts(x) \cap l' = \emptyset$. Now $\gamma \models_l pts$ amounts to $\gamma \models_{l'} pts$. Hence we get $\gamma' \models_{l'} pts'$ because $z := e : (pts, l') \rightarrow (pts', l')$ and by soundness of $(:=_1^l)$.
- (f) The type derivation has the form $(*_2^l)$. In this case, for every $z' \in pts(x)$, we have $z := e : pts \rightarrow pts', \gamma(x) = z'$, $z := e : \gamma \rightarrow \gamma'$, and $l = l' \cup FV(e) \cup \{x\}$. We have $z \in pts(x)$ because $\gamma \models_l pts$ and $x \in l$. Therefore by $(:=_2^l)$ we have $x := z : (pts, (l' \setminus \{z\}) \cup FV(e)) \rightarrow (pts', l')$.

$\gamma \models_l pts$ implies $\gamma \models_{(l' \setminus \{z\}) \cup FV(e)} pts$. Hence by soundness of $(:=_2^l)$, we get $\gamma' \models_{l'} pts'$.

- (g) The type derivation has the form (par^l) . In this case there exist a permutation $\theta : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ and $n + 1$ states $\gamma = \gamma_1, \dots, \gamma_{n+1} = \gamma'$ such that for every $1 \leq i \leq n$, $S_{\theta(i)} : \gamma_i \rightarrow \gamma_{i+1}$. Also $\gamma_1 \models_l pts$ implies $\gamma_1 \models_{l_{\theta(1)}} pts \cup \bigcup_{j \neq \theta(1)} pts_j$. Therefore by the induction hypothesis $\gamma_2 \models_{l' \cup \bigcup_{j \neq \theta(1)} l_j} pts_{\theta(1)}$. This implies $\gamma_2 \models_{l_{\theta(2)}} pts \cup \bigcup_{j \neq \theta(2)} pts_j$. Again by the induction hypothesis we get $\gamma_3 \models_{l' \cup \bigcup_{j \neq \theta(2)} l_j} pts_{\theta(2)}$. Therefore by a simple induction on n , we can show that $\gamma' = \gamma_{n+1} \models_{l' \cup \bigcup_{j \neq \theta(n)} l_j} pts_{\theta(n)}$ which implies $\gamma' \models_{l'} pts' = \bigcup_j pts_j$.

- (h) The type derivation has the form $(par\text{-}for^l)$: In this case there exists n

$\overbrace{\text{such that } par\{\{S\}, \dots, \{S\}\} : \gamma \rightsquigarrow \gamma'}^{n\text{-times}}$

$S : (pts \cup pts', l) \rightarrow (pts', l \cup l')$. By (par^l) we conclude that $par\{\{S\}, \dots, \{S\}\} : \overbrace{(pts, l) \rightarrow (pts', l')}^{n\text{-times}}$. Therefore by soundness of (par^l) , we get $\gamma' \models_{l'} pts'$.

3. The proof is also by induction on the structure of type derivation and it is straightforward.

The proof of the following corollary follows from Theorem 2.

Corollary 1. Suppose $S : \gamma \rightsquigarrow \gamma'$ and $S : (pts, l) \rightarrow (pts', l')$. Then $\gamma \models (pts, l)$ implies $\gamma' \models (pts', l')$.

Theorem 3. Suppose that $S : (pts, l) \rightarrow (pts', l')$, $S : \gamma \rightsquigarrow \gamma'$, $\gamma \sim_{(pts, l)} \gamma_*$, and S does not abort at γ_* . Then there exists a state γ'_* such that $S : \gamma_* \rightarrow \gamma'_*$ and $\gamma' \sim_{(pts', l')} \gamma'_*$.

Proof. The proof is by induction on structure of type derivation. We demonstrate some cases:

1. The type derivation has one of the forms $(:=_1^l)$ and $(:=_2^l)$. In this case, $pts' = pts[x \mapsto A]$ and $\gamma' = \gamma[x \mapsto \llbracket e \rrbracket \gamma]$. We take $\gamma'_* = \gamma_*[x \mapsto \llbracket e \rrbracket \gamma_*]$.
2. The type derivation has the form $(:=_1^{*l})$ or $(:=_2^{*l})$. In this case, $\forall z' \in pts(y)$, we have $x := z : pts \rightarrow pts', \gamma(y) = z'$, and $x := z : \gamma \rightarrow \gamma'$. We set $\gamma'_* = \gamma_*[x \mapsto \gamma_*(z)]$.
3. The type derivation has one of the forms $(:=_1^{*l})$ and $(:=_2^{*l})$. In this case, $\forall z' \in pts(x)$, we have $z := e : pts \rightarrow pts', \gamma(x) = z'$, and $z := e : \gamma \rightarrow \gamma'$. We let $\gamma'_* = \gamma_*[z \mapsto \llbracket e \rrbracket \gamma_*]$.
4. The type derivation has the form (par^l) . In this case there exist a permutation $\theta : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ and $n + 1$ states $\gamma = \gamma_1, \dots, \gamma_{n+1} = \gamma'$ such that for every $1 \leq i \leq n$, $S_{\theta(i)} : \gamma_i \rightarrow \gamma_{i+1}$. We refer to γ_* as γ_{*1} . We have $\gamma_1 \sim_{(pts, \bigcup_{l_i} l_i)} \gamma_{*1}$ which implies $\gamma_1 \sim_{(pts \cup \bigcup_{j \neq \theta(1)} pts_j, l_{\theta(1)})} \gamma_{*1}$. Therefore by induction hypothesis, there exists γ_{*2} such that $S_{\theta(1)} : \gamma_{*1} \rightarrow \gamma_{*2}$ and $\gamma_2 \sim_{(pts_{\theta(1)}, l' \cup \bigcup_{j \neq \theta(1)} l_j)} \gamma_{*2}$ which implies $\gamma_2 \sim_{(pts \cup \bigcup_{j \neq \theta(2)} pts_j, l_{\theta(2)})} \gamma_{*2}$. Therefore a simple induction on n proves the required.

5 Dead code elimination

This section introduces a type system for dead code elimination. Given a program and a set of variables whose values concern us at the end of the program, there may be some code in the program that has no effect on the values of these variables. Such code is called *dead code*. The type system presented here aims at optimizing structured parallel programs with pointer constructs via eliminating dead code. In the form of a type derivation, the type system associates each optimization with a proof for the soundness of the optimization. Optimizing a program may result in correcting it i.e. preventing it from aborting. Of course this happens if the removed dead code is the only cause of abortion.

The type system presented here has judgements of the form: $S : (pts, l) \rightarrow (pts', l') \hookrightarrow S'$. The intuition is that S' optimizes S towards dead code elimination (and may be program correction). As mentioned early in many occasions, the derivation of such judgement provides a justification for the optimization process. The form of the judgement makes it apparent that the type system presented in this section is built on the type system for live-variables analysis.

Algorithm: *parallel-optimize*

- Input : a statement S of the language presented in Section 2 and a set of variables l' that we consider live (their values concern us) at the end of executing S ;
- Output: an optimized and may be corrected version S' of S such that the relation between S and S' is as stated in Theorem 4.
- Method :
 1. Find pts such that $S : \perp \rightarrow pts$ in the type system for pointer analysis.
 2. Find l such that $S : (\perp, l) \rightarrow (pts, l')$ in the type system for live-variables analysis.
 3. Find S' such that $S : (\perp, l) \rightarrow (pts, l') \hookrightarrow S'$ in the type system for dead code elimination.

Fig. 3. The algorithm *optimize-parallel*

Figure 3 outlines an algorithm, *parallel-optimize*, that summarizes the optimization process. A pointer analysis that annotates the points of the input program with pointer information is the first step of the algorithm. This step takes the form of a post type derivation of S , in our type system for pointer analysis, using the bottom points-to type $\perp = \{x \mapsto \emptyset \mid x \in Var\}$ as the pre type. Secondly, the algorithm refines the pointer information obtained in the first step via annotating the pointer types with type components for live-variables. Using our type systems for live-variables analysis, this is done via a pre type derivation of S for the set l' , the set of variables whose values concerns us at the end of execution, as the post type. Finally, the information obtained so far is utilized in the third step to find S' via using the type system for dead code elimination proposed in this section. Applying this algorithm to the program on the left-hand side of Figure 1 results in the program on the right-hand side of the same figure. The details of this application is a simple exercise.

The inference rules of our type system for dead code elimination are as follows:

$$\begin{array}{c}
\frac{x := e : pts \rightarrow pts' \quad x \notin l'}{x := e : (pts, l') \rightarrow (pts', l') \hookrightarrow skip} (:=_1^e) \\
\frac{x := e : pts \rightarrow pts' \quad x \in l'}{x := e : (pts, (l' \setminus \{x\}) \cup FV(e)) \rightarrow (pts', l') \hookrightarrow x := e} (:=_2^e) \\
\frac{x \notin l'}{x := \&y : (pts, l') \rightarrow (pts[x \mapsto \{y'\}], l') \hookrightarrow skip} (:= \&_1^e) \quad \frac{}{skip : (pts, l) \rightarrow (pts, l) \hookrightarrow skip} \\
\frac{x \in l'}{x := \&y : (pts, l' \setminus \{x\}) \rightarrow (pts[x \mapsto \{y'\}], l') \hookrightarrow x := \&y} (:= \&_2^e) \\
\frac{x := *y : pts \rightarrow pts' \quad x \notin l'}{x := *y : (pts, l' \cup \{y\}) \rightarrow (pts', l') \hookrightarrow skip} (:= *_1^e) \\
\frac{x := *y : pts \rightarrow pts' \quad x \in l'}{x := *y : (pts, (l' \setminus \{x\}) \cup \{y, z \mid z' \in pts(y)\}) \rightarrow (pts', l') \hookrightarrow x := *y} (:= *_2^e) \\
\frac{*x := e : pts \rightarrow pts' \quad pts(x) \cap l = \emptyset}{*x := e : (pts, l' \cup \{x\}) \rightarrow (pts', l') \hookrightarrow skip} (* :=_1^e) \\
\frac{*x := e : pts \rightarrow pts' \quad pts(x) \cap l' \neq \emptyset}{*x := e : (pts, l' \cup \{x\} \cup FV(e)) \rightarrow (pts', l') \hookrightarrow *x := e} (* :=_1^e) \\
\frac{S_i : (pts \cup \cup_{j \neq i} pts_j, l_i) \rightarrow (pts_i, l' \cup \cup_{j \neq i} l_j) \hookrightarrow S'_i}{par\{\{S_1\}, \dots, \{S_n\}\} : (pts, \cup_i l_i) \rightarrow (\cup_i pts_i, l') \hookrightarrow par\{\{S'_1\}, \dots, \{S'_n\}\}} (par^e) \\
\frac{par\{\{if\ b_1\ then\ S_1\ else\ skip\}, \dots, \{if\ b_n\ then\ S_n\ else\ skip\}\} : (pts, l) \rightarrow (pts', l') \hookrightarrow par\{\{if\ b_1\ then\ S'_1\ else\ skip\}, \dots, \{if\ b_n\ then\ S'_n\ else\ skip\}\}}{par-if\{(b_1, S_1), \dots, (b_n, S_n)\} : (pts, l) \rightarrow (pts', l') \hookrightarrow par-if\{(b_1, S'_1), \dots, (b_n, S'_n)\}} (par-if^e) \\
\frac{S : (pts \cup pts', l) \rightarrow (pts', l' \cup l) \hookrightarrow S'}{par-for\{S\} : (pts, l) \rightarrow (pts', l') \hookrightarrow par-for\{S'\}} (par-for^e) \\
\frac{S_1 : (pts, l) \rightarrow (pts'', l'') \hookrightarrow S'_1 \quad S_2 : (pts'', l'') \rightarrow (pts', l') \hookrightarrow S'_2}{S_1; S_2 : (pts, l) \rightarrow (pts', l') \hookrightarrow S'_1; S'_2} (seq^e) \\
\frac{S_i : (pts, l) \rightarrow (pts', l') \hookrightarrow S'_i \quad S_f : (pts, l) \rightarrow (pts', l') \hookrightarrow S'_f}{if\ b\ then\ S_i\ else\ S_f : (pts, l \cup FV(b)) \rightarrow (pts', l') \hookrightarrow if\ b\ then\ S'_i\ else\ S'_f} (if^e) \\
\frac{l = l' \cup FV(b) \quad S_i : (pts, l') \rightarrow (pts, l) \hookrightarrow S'_i}{while\ b\ do\ S_i : (pts, l) \rightarrow (pts, l') \hookrightarrow while\ b\ do\ S'_i} (whl^e) \\
\frac{(pts'_1, l'_1) \leq (pts_1, l_1) \quad S : (pts_1, l_1) \rightarrow (pts_2, l_2) \hookrightarrow S' \quad (pts_2, l_2) \leq (pts'_2, l'_2)}{S : (pts'_1, l'_1) \rightarrow (pts'_2, l'_2) \hookrightarrow S'} (csq^e)
\end{array}$$

When optimizing programs it is important to guarantee that if (a) the original and optimized programs are executed in similar states, and (b) the original program ends at a state (rather than *abort*), then (a) the optimized program does

not abort as well, and (b) the optimized program reaches a state similar to that reached by the original program. Indeed, this is guaranteed by the following theorem.

Theorem 4. (Soundness) Suppose that $S : (pts, l) \rightarrow (pts', l') \hookrightarrow S'$ and $\gamma \sim_{(pts, l)} \gamma_*$. Then

1. If $S : \gamma \rightsquigarrow \gamma'$, then there exists a state γ'_* such that $S' : \gamma_* \rightarrow \gamma'_*$ and $\gamma' \sim_{(pts', l')} \gamma'_*$.
2. If $S' : \gamma_* \rightarrow \gamma'_*$ and S does not abort at γ , then there exists a state γ' such that $S : \gamma \rightsquigarrow \gamma'$ and $\gamma' \sim_{(pts', l')} \gamma'_*$.

The proof of this theorem is by induction on the structure of type derivation and it follows smoothly from Theorem 3. More precisely Theorem 3 is used when $S' = S$. When $S' = skip$, we take $\gamma'_* = \gamma_*$ in 1. We note that the requirement of Theorem 3 that S does not abort at γ_* is guaranteed when this theorem is called in the proof of Theorem 4.

6 Related work

Analysis of multithreaded programs: The analysis of multithreaded programs is an area that receives growing interest. It is a challenging area [5] as the presence of threading complicates the program analysis. The work in this area can be classified into two main categories. One category includes techniques that were designed specifically to optimize or correct multithreaded programs. The other category includes techniques whose scope was extended from sequential programs to multithreaded programs.

The first category mentioned above covers several directions of research; synchronization analysis, deadlock, data race, and memory consistency. The purpose in the analysis of synchronization constructs [6,7] is to clarify how the synchronization actions apart executions of program segments. The result of this analysis can be used by compiler to conveniently add join-fork constructs. Clearly, adding such join-fork constructs will reduce the run time of the program. One problem of multithreading computing is deadlock which results from round waiting to gain resources. Researchers have developed various techniques for deadlock detection [8,9,10]. The situation when a memory location is accessed by two threads (one of them writes in the location) without synchronization is called data race. On direction of research in this category focuses on data race detection [11]. The analysis of multithreaded programs becomes even harder in the presence of a weak memory consistency model because such model does not guarantee that a write statement included in one thread is observed by other threads in the same order. However such model simplifies some issues on the hardware level. The work in this direction, like [12], aims at overcomes the drawbacks of using a simple consistency memory model. Asymmetric Distributed Shared Memory (ADSM), a programming model serving heterogeneous computing, is introduced in [12]. ADSM manages a shared virtual memory to enable CPUs-access to addresses on the accelerator real memory.

Under the second category mentioned above comes several directions of research. One such direction is the using of flow-insensitive analysis techniques to analyze multithreaded programs [13,14]. Although flow-insensitive techniques are not very precise, some applications can afford that. Examples of program analyses whose techniques were extended to cover multithreaded programs are code motion [15], constant propagation [16], data flow for multithreaded programs with copy-in and copy-out memory semantics [17,18], and concurrent static single assignment form [19].

The problem with almost all the work referred to above is that it does not apply to pointer programs. More precisely, for some of the work the application is possible only if we have the result of a pointer analysis for the input pointer program. The technique presented in this paper for optimizing multithreaded programs has the advantage of being simpler and more reliable than the optimization techniques referred to above that would work in the presence of a pointer analysis.

Pointer analysis: The pointer analysis for sequential programs has been studied extensively for decades [20]. One way to classify the work in this area is according to properties of flow-sensitivity and context-sensitivity. Hence the work is classified into flow-sensitivity, flow-insensitivity, context-sensitivity, and context-insensitivity.

Flow-sensitive analyses [21,22,23], which are more natural than flow-insensitive to most applications, consider the order of program commands. Mostly these analyses perform an abstract interpretation of program using dataflow analysis to associate each program point with a points-to relation. Flow-insensitive analyses [24,25] do not consider the order of program commands. Typically the output of these analyses, which are performed using a constraint-based approach, is a points-to relation that is valid all over the program. Clearly the flow-sensitive approach is more precise but less efficient than the flow-insensitive one. Moreover flow-insensitive techniques can be used to analyze multithreaded programs.

The idea of context-sensitive approach [26,23] is to produce a points-to relation for the context of each call site of each procedure. On the other hand, the context-insensitive [27] pointer analysis produces one points-to relation for each procedure to cover contexts of all call sites. As expected the context-sensitive approach is more precise but less efficient than the context-insensitive one.

Although the problem of pointer analysis for sequential programs was studied extensively, a little effort was done towards a pointer analysis for multithreaded programs. In [28], a flow sensitive analysis for multithreaded programs was introduced. This analysis associates each program point with a triple of points-to relations. This in turn complicates the analysis and creates a sort of redundancy in the collected points-to information. Investigating the details of this approach and our work makes it apparent that our work is simpler and more accurate than this approach. Moreover our approach provides a proof for the correctness of the pointer analysis for each program. To the best of

our knowledge, such proof is not known to be provided by any other existing approach.

Type systems in program analysis: The work in [1,2,3,29,30,31] is among the closest work to ours in the sense that it uses type systems to achieve the program analysis in a way similar to the present paper. The work in [32] can be seen as a special case of our work for the case of while language where there is no threading nor pointer constructs.

The work in [2] shows that a good deal of program analysis can be done using type systems. More precisely, it proves that for every analysis in a certain class of data-flow analyses, there exists a type system such that a program checks with a type if and only if the type is a supertype for the set resulting from running the analysis on the program. The type system in [33] and the flow-logic work in [3], which is used in [34] to study security of the coordinated systems, are very similar to [2]. Moreover, the work [33] transforms logical statements about programs to statements about the program optimizations. For the simple while language, the work in [1] introduces type systems for constant folding and dead code elimination and also logically proves correctness of optimizations. The bidirectional data-flow analyses and their program optimizations are treated with type systems in [35]. Earlier, related work (with structurally-complex type systems) is [36]. The work in [30] presents type systems that checks memory safety of multithreaded programs using sensitive-nonsensitive pointer analysis.

To the best of our knowledge, our approach is the first attempt to use type systems to optimize multithreaded programs and associates every individual optimization with a justification for correctness.

References

1. Nick Benton. *Simple relational correctness proofs for static analyses and program transformations*. In Neil D. Jones and Xavier Leroy, editors, POPL, pages 14–25. ACM, 2004.
2. Peeter Laud, Tarmo Uustalu, and Varmo Vene. *Type systems equivalent to data-flow analyses for imperative languages*. Theor. Comput. Sci., 364(3):292–310, 2006.
3. Hanne Riis Nielson and Flemming Nielson. *Flow logic: A multi-paradigmatic approach to static analysis*. In Torben Mogensen, David A. Schmidt, and Ivan Hal Sudborough, editors, The Essence of Computation, volume 2566 of Lecture Notes in Computer Science, pages 223–244. Springer, 2002.
4. C. A. R. Hoare. *An axiomatic basis for computer programming*. Commun. ACM, 12(10):576–580, 1969.
5. Vivek Sarkar. *Challenges in code optimization of parallel programs*. In Oege de Moor and Michael I. Schwartzbach, editors, CC, volume 5501 of Lecture Notes in Computer Science, page 1. Springer, 2009.
6. Chen Tian, Vijay Nagarajan, Rajiv Gupta, and Sriraman Tallam. *Automated dynamic detection of busy-wait synchronizations*. Softw., Pract. Exper., 39(11):947–972, 2009.
7. Chuanfu Xu, Yonggang Che, Jianbin Fang, and Zhenghua Wang. *Optimizing adaptive synchronization in parallel simulators for large-scale parallel systems and applications*. In CIT, pages 131–138. IEEE Computer Society, 2010.

8. Byung-Chul Kim, Sang-Woo Jun, Dae Joon Hwang, and Yong-Kee Jun. *Visualizing potential deadlocks in multithreaded programs*. In Victor Malyshev, editor, PaCT, volume 5698 of Lecture Notes in Computer Science, pages 321–330. Springer, 2009.
9. Yin Wang, Terence Kelly, Manjunath Kudlur, Steephane Lafortune, and Scott A. Mahlke. *Gadara: Dynamic deadlock avoidance for multithreaded programs*. In Richard Draves and Robbert van Renesse, editors, OSDI, pages 281–294. USENIX Association, 2008.
10. Xiang Xiao and Jaehwan John Lee. *A true $o(1)$ parallel deadlock detection algorithm for single-unit resource systems and its hardware implementation*. IEEE Trans. Parallel Distrib. Syst., 21(1):4–19, 2010.
11. K. Leung, Zhiyi Huang, Qihang Huang, and Paul Werstein. *Maotai 2.0: Data race prevention in view-oriented parallel programming*. In PDCAT, pages 263–271. IEEE Computer Society, 2009.
12. Isaac Gelado, Javier Cabezas, Nacho Navarro, John E. Stone, Sanjay J. Patel, and Wen mei W. Hwu. *An asymmetric distributed shared memory model for heterogeneous parallel systems*. In James C. Hoe and Vikram S. Adve, editors, ASPLOS, pages 347–358. ACM, 2010.
13. Enno Lubbers and Marco Platzner. *Communication and synchronization in multithreaded reconfigurable computing systems*. In Toomas P. Plaks, editor, ERSAC, pages 83–89. CSREA Press, 2008.
14. Erik Ruf. *Effective synchronization removal for java*. In PLDI, pages 208–218, 2000.
15. Jens Knoop and Bernhard Steffen. *Code motion for explicitly parallel programs*. In PPOPP, pages 13–24, 1999.
16. Jaejin Lee, Samuel P. Midkiff, and David A. Padua. *A constant propagation algorithm for explicitly parallel programs*. International Journal of Parallel Programming, 26(5):563–589, 1998.
17. Johnson S. Kin and Jose Luis Pino. *Multithreaded synchronous data flow simulation*. In DATE, pages 11094–11095. IEEE Computer Society, 2003.
18. Mengxiao Liu, Weixing Ji, Xing Pu, and Jiaxin Li. *A parallel memory system model for multi-core processor*. In NAS, pages 219–222. IEEE Computer Society, 2009.
19. Jaejin Lee, Samuel P. Midkiff, and David A. Padua. *Concurrent static single assignment form and constant propagation for explicitly parallel programs*. In Zhiyuan Li, Pen-Chung Yew, Siddhartha Chatterjee, Chua-Huang Huang, P. Sadayappan, and David C. Sehr, editors, LCPC, volume 1366 of Lecture Notes in Computer Science, pages 114–130. Springer, 1997.
20. Michael Hind. *Pointer analysis: haven't we solved this problem yet?* In PASTE, pages 54–61, 2001.
21. Ben Hardekopf and Calvin Lin. *Semi-sparse flow-sensitive pointer analysis*. In Zhong Shao and Benjamin C. Pierce, editors, POPL, pages 226–238. ACM, 2009.
22. Ji Wang, Xiaodong Ma, Wei Dong, Hou-Feng Xu, and Wanwei Liu. *Demand-driven memory leak detection based on flow- and context-sensitive pointer analysis*. J. Comput. Sci. Technol., 24(2):347–356, 2009.
23. Hongtao Yu, Jingling Xue, Wei Huo, Xiaobing Feng 0002, and Zhaoqing Zhang. *Level by level: making flow- and context-sensitive pointer analysis scalable for millions of lines of code*. In Andreas Moshovos, J. Gregory Steffan, Kim M. Hazelwood, and David R. Kaeli, editors, CGO, pages 218–229. ACM, 2010.
24. Stephen Adams, Thomas Ball, Manuvir Das, Sorin Lerner, Sriram K. Rajamani, Mark Seigle, and Westley Weimer. *Speeding up dataflow analysis using flow-insensitive pointer analysis*. In Manuel V. Hermenegildo and German Puebla, editors, SAS, volume 2477 of Lecture Notes in Computer Science, pages 230–246. Springer, 2002.

25. Paul Anderson, David Binkley, Genevieve Rosay, and Tim Teitelbaum. *Flow insensitive points-to sets*. Information & Software Technology, 44(13):743–754, 2002.
26. Rupesh Nasre, Kaushik Rajan, Ramaswamy Govindarajan, and Uday P. Khedker. *Scalable context-sensitive points-to analysis using multi-dimensional bloom filters*. In Zhenjiang Hu, editor, APLAS, volume 5904 of Lecture Notes in Computer Science, pages 47–62. Springer, 2009.
27. Donglin Liang, Maikel Pennings, and Mary Jean Harrold. *Extending and evaluating flow-insensitive and context-insensitive points-to analyses for java*. In PASTE, pages 73–79, 2001.
28. Radu Rugina and Martin C. Rinard. *Pointer analysis for structured parallel programs*. ACM Trans. Program. Lang. Syst., 25(1):70–116, 2003.
29. Mohamed A. El-Zawawy. *Program optimization based pointer analysis and live stack-heap analysis*. International Journal of Computer Science Issues, 8(2), March 2011.
30. Mohamed A. El-Zawawy. *Flow sensitive-insensitive pointer analysis based memory safety for multithreaded programs*. In Beniamino Murgante, Osvaldo Gervasi, Andres Iglesias, David Taniar, and Bernady O. Apduhan, editors, ICCSA (5), volume 6786 of Lecture Notes in Computer Science, pages 355–369. Springer, 2011.
31. Mohamed A. El-Zawawy. *Probabilistic pointer analysis for multithreaded programs*. Accepted for publication in ScienceAsia.
32. A. Saabas and T. Uustalu. *Program and proof optimizations with type systems*. The 16th Nordic Workshop on the Programming Theory (NWPT 2006). Journal of Logic and Algebraic Programming, 77(1-2):131 – 154, 2008.
33. Mayur Naik and Jens Palsberg. *A type system equivalent to a model checker*. ACM Trans. Program. Lang. Syst., 30(5):1–24, 2008.
34. Rocco De Nicola, Daniele Gorla, Rene Rydhof Hansen, Flemming Nielson, Hanne Riis Nielson, Christian W. Probst, and Rosario Pugliese. *From flow logic to static type systems for coordination languages*. Sci. Comput. Program., 75(6):376– 397, 2010.
35. Maria Frade, Ando Saabas, and Tarmo Uustalu. *Bidirectional data-flow analyses, type-systematically*. In German Puebla and German Vidal, editors, PEPM, pages 141–150. ACM, 2009.
36. Jens Palsberg and Patrick O’Keefe. *A type system equivalent to flow analysis*. ACM Trans. Program. Lang. Syst., 17(4):576–599, 1995.